International Journal of Psychophysiology 83 (2012) 56–64

Contents lists available at SciVerse ScienceDirect

International Journal of Psychophysiology

journal homepage: www.elsevier.com/locate/ijpsycho

# Memory strength and specificity revealed by pupillometry

Megan H. Papesh *, Stephen D. Goldinger, Michael C. Hout

*Arizona State University, United States*

ARTICLE INFO

*Article history:*
Received 16 June 2011
Received in revised form 30 September 2011
Accepted 4 October 2011
Available online 20 October 2011

*Keywords:*
Recognition memory
Specificity effects
Pupillometry

ABSTRACT

Voice-specificity effects in recognition memory were investigated using both behavioral data and pupillometry. Volunteers initially heard spoken words and nonwords in two voices; they later provided confidence-based old/new classifications to items presented in their original voices, changed (but familiar) voices, or entirely new voices. Recognition was more accurate for old-voice items, replicating prior research. Pupillometry was used to gauge cognitive demand during both encoding and testing: enlarged pupils revealed that participants devoted greater effort to encoding items that were subsequently recognized. Further, pupil responses were sensitive to the cue match between encoding and retrieval voices, as well as memory strength. Strong memories, and those with the closest encoding-retrieval voice matches, resulted in the highest peak pupil diameters. The results are discussed with respect to episodic memory models and Whittlesea's (1997) SCAPE framework for recognition memory.

## 1. Introduction

In the present study, we examined the extent to which memory strength and specificity for spoken items are revealed by pupillometry across learning and recognition. Although the speech signal is characterized by idiosyncratic variations that listeners must fluently overcome, debate surrounds the necessity of *encoding* this information into memory during on-line perception. For example, people encounter little perceptual resistance when processing the same words spoken by different speakers, each of whom has a unique vocal structure, pattern of intonation, and speaking rate. Changes in context and other non-linguistic variables similarly pose little challenge to the perceptual system. Speech perception is clearly robust to idiosyncratic variations in the input signal, but do these variations get “filtered out” during perception, or are they somehow stored in a detailed memory trace, capable of affecting subsequent perception or retrieval?

Two general, and opposing, approaches to this problem have dominated the literature. According to the first, the speech signal is stripped of idiosyncratic information upon encoding, allowing the perceiver to activate abstract representations in memory (Joos, 1948). Such theories generally treat idiosyncratic variations as undesirable noise in the speech signal, a problem for the perceptual system to overcome (Pisoni, 1993). According to the second approach, surface properties of speech are stored in unique, episodic traces; surface information is not noise, but is instead utilized to aid subsequent recognition (McLennan and Luce, 2005). These theoretical approaches are denoted *abstractionist* and *episodic* theories, respectively. Both views have empirical support, and either would allow listeners to resolve variability in speech, yielding immediate and effortless mapping of speech signals to segmental and lexical representations.

Abstractionist theories posit *normalization* mechanisms that “correct” the speech signal for its idiosyncratic properties, allowing perception to operate at the level of stored, ideal representations. Normalization is proposed to explain the constancy of linguistic perception, despite variations across talkers and contexts (see Bradlow et al., 1999; Magnuson and Nusbaum, 2007). For example, Marslen-Wilson and Warren (1994; Lahiri and Marslen-Wilson, 1991) proposed an account whereby variations in spoken words are immediately resolved; lexical access occurs when a stored lexical unit receives sufficient activation from this corrected input (see McClelland and Elman, 1986). Attention is directed to the level of word meaning (as it should be); later, memory for surface information should be negligible. Abstractionist theories are both logically appealing (see Bowers, 2000; McQueen et al., 2006) and have empirical support. For example, early in word perception, priming effects appear to be mediated by abstract phonemic representations, devoid of superficial details (McLennan and Luce, 2005; see also McQueen et al., 2003).

Although the normalizing process has intuitive appeal, there are many empirical demonstrations that spoken word perception creates detailed, episodic memory traces. In theoretical terms, episodic theories have many desirable properties. For example, the correctly predict that speech perception becomes more robust as people are exposed to a wider range of exemplars (Lively et al., 1993). Of greater importance, they provide a natural mechanism to explain *specificity effects* in perception and memory (Goldinger, 1996; McLennan and Luce, 2005). Across many experimental paradigms, performance is affected (usually improved) when items presented in a study phase

---

* Corresponding author at: Department of Psychology, Box 871104, Arizona State University, Tempe, AZ 85287-1104, United States. Tel.: +1 480 965 1377; fax: +1 480 965 8544.

*E-mail address:* megan.papesh@asu.edu (M.H. Papesh).

0167-8760/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.ijpsycho.2011.10.002

are later repeated in a test phase. Such improvements are typically stronger when surface information (such as the speaker's voice) is preserved across study and test. These *voice specificity effects* are especially robust in implicit memory. For example, when the voice associated with a word changes across study and test, it reduces performance in perceptual identification (Church and Schacter, 1994; Goldinger, 1996; Pilotti et al., 2000; Sheffert, 1998), naming (Goldinger, 1998), lexical decision (Luce and Lyons, 1998), and word-stem completion (Church and Schacter, 1994; Schacter and Church, 1992, but see Pilotti et al., 2000). Exposure to voice-specific tokens of words can also affect later speech production (Goldinger, 1998; Goldinger and Azuma, 2004; Pardo, 2006; Shockley et al., 2004).

In tests of explicit memory, word recognition is again typically best for items that are tested in their original study voices. For example, Palmeri et al. (1993; Sheffert and Fowler, 1995; Senkfor and Van Petten, 1998) observed benefits for same-voice repetitions in continuous recognition memory, relative to changed-voice repetitions. Voice effects were robust for lists including as many as 20 talkers and study-test lags up to 64 items. Goldinger (1996) further examined voice effects, manipulating the similarity among talkers, levels of initial processing, and extending the retention time out to a week. He found same-voice advantages in both implicit and explicit measures, although the same-voice advantage vanished in explicit recognition after a 1-week delay (see also Goh, 2005). In the present research, we further examined voice effects in recognition memory, complementing the standard paradigm with real-time measures of changes in pupil diameters. By doing so, we were able to assess participants' cognitive effort during word encoding and test, with special emphasis on potential voice effects: If people encode voice-specific traces of spoken words, later repetition of those same tokens should reduce cognitive effort, as indicated by the pupillary reflex.

The present paradigm also allowed us to assess whether voice changes across study and test influence memory strength. By obtaining 1–6 confidence estimates alongside recognition judgments (e.g., 1 = *very sure new*; 6 = *very sure old*), we examined subjective memory strength in two ways. First, confidence estimates allowed us to create receiver operating characteristic (ROC) curves, plotting the hit and false-alarm rates at various levels of confidence or bias (Macmillan & Creelman, 2005). ROC curves are also commonly $z$-transformed and plotted on standardized axes ($z$-ROCs). Different memory theories make different predictions regarding the shapes of these curves (Wixted, 2007; Yonelinas and Parks, 2007). For example, *dual-process* theories (e.g., Yonelinas, 1994) assume that two processes subserve memory decisions: A thresholded, all-or-none recollection process and a graded, strength-based familiarity process. These theories generally predict that ROC curves will be linear and $z$-ROCs will be curvilinear. On the other hand, *strength-based* theories (Wixted and Mickes, 2010) assume that recollection and familiarity signals are both graded and summed into a single memory strength signal, which is used as the basis for recognition decisions. Strength theories generally predict curvilinear ROCs and linear $z$-ROCs. In the present study, we collected confidence estimates, allowing us to assess the underlying strength distributions for targets and lures. We also examined pupillary reflexes during study as a function of *subsequent* confidence. If pupil dilation reflects part of the recognition memory process, as suggested by Võ et al. (2008) and Kafkas and Montaldi (2011), then pupillary changes during encoding may accurately "track" subsequent estimates of memory strength. This approach allowed us to assess whether pupillometry reveals differences in strong versus weak memories across both encoding and retrieval.

## 1. Pupillometry

Pupillary reflexes occur during all forms of visual and cognitive processing, and are hypothesized to reflect brain activity during processing (Beatty and Kahneman, 1966). Enlarged pupils are typically associated with increased cognitive demand (Porter et al., 2007) and provide sensitive indices of cognitive effort, similar to ERP waveforms (Beatty, 1982). Using the *subsequent memory paradigm*, researchers can compare neurophysiological measures across study and test to differentiate the neural activity associated with subsequently remembered versus forgotten information. Such investigations have been reported using fMRI (e.g., Ranganath et al., 2005) and ERP (Cansino and Trejo-Morales, 2008; Duarte et al., 2004; Guo et al., 2006). In the current investigation, we used pupillometry to compare the effort involved in cognitive operations across study and test, with emphasis on voice effects. We had two key questions: First, does greater effort during encoding predict greater success during recognition? And second, does preservation of voice information influence processing during test?

The appeal of pupillometry to the investigation of cognitive phenomena lies in its automaticity. Pupillary reflexes are controlled by the sympathetic and parasympathetic systems, which hold reciprocal connections to central nervous system (CNS), suggesting that they may exert an influence on CNS structures relevant to cognition (Gianaros et al., 2004). Pupils dilate following sympathetic system activation and/or parasympathetic system inhibition and constrict following activity of the parasympathetic system (Steinhauer et al., 2004). Although the pupils change reflexively in response to general factors, such as emotional arousal and anxiety, such *tonic* changes are independent of *phasic* changes, which arise upon the onset of stimuli for cognitive processing. Such phasic changes are known as task-evoked pupillary responses (TEPRs), and have long been used to infer cognitive effort across domains such as lexical decision (Kuchinke et al., 2007), attention allocation (Karatekin et al., 2004), working memory load (Granholm et al., 1996; Van Gerven et al., 2004), face perception (Goldinger et al., 2009), and general cognitive processing (Granholm and Verney, 2004). In fact, Kahneman (1973) used TEPRs as his primary index of mental processing load in his theory of attention, owing to its sensitivity to variations within or between tasks, and its ability to reflect individual differences in cognitive ability.

The relationship between human memory and the pupillary reflex has seldom been investigated, but animal models suggest a potential relationship between pupil dilation and memory encoding/retrieval (Croiset et al., 2000). As has been shown with rats (Clark et al., 1995), stimulating the vagus nerve in the parasympathetic pathway in humans (patients undergoing treatment for epilepsy) enhances memory retention, if stimulation is applied during consolidation (Clark et al., 1999). Such findings suggest a modulatory influence of autonomic activity on memory formation and retrieval. The first study examining human memory and the pupillary reflex was reported by Võ et al. (2008), who noted the similarity between pupillary and ERP waveforms, which are known to reflect memorial processes (Dietrich et al., 2000; Johansson et al., 2004). Võ et al. observed a "pupillary old/new effect," wherein pupils were larger during study trials leading to hits, relative to correct rejections. They interpreted this effect in a dual-process framework (Yonelinas, 2001, 2002), suggesting that enlarged pupils were observed for hits because they included recollection, which is hypothesized to be a slow, cognitively demanding process. Similar effects were reported by Papesh and Goldinger (2011), who found a pupillary old/new effect across study and test presentations of auditory low- and high-frequency words. Specifically, when participants studied words that were subsequently remembered, those trials were associated with enlarged pupils, relative to subsequently forgotten and new words. This pattern was especially strong for low-frequency words, suggesting that memorial encoding, coupled with the cognitive operations usurped in processing low-frequency words (see Goldinger and Papesh, 2009; Kuchinke et al., 2007; Papesh and Goldinger, 2008), resulted in an overall increase in cognitive demand. In prior studies (Goldinger, 1996, 1998), voice effects were stronger for low-frequency words, relative to high-frequency words, and were stronger for nonwords, relative to words.